

Personalized Prediction of Driving Energy Consumption based on Participatory Sensing

Chien-Ming Tseng and Chi-Kin Chau

Abstract—The advent of abundant on-board sensors and electronic devices in vehicles populates the paradigm of participatory sensing for intelligent transportation systems, by harnessing crowd-sourced data gathering to enable knowledge discovery in various applications, such as distance-to-empty prediction and eco-routing. While participatory sensing can provide various driving and vehicle data, there lacks a systematic study on effectively utilizing the data for personalized prediction applications. There are considerable challenges on how to interpolate the missing data from a sparse dataset, which often arises from participatory sensing. This paper presents and compares various personalized prediction approaches for driving energy consumption, including a blackbox approach that identifies driver/vehicle/environment-dependent factors and a collaborative filtering approach that uses matrix factorization. To evaluate the effectiveness of our approaches, a case study of distance-to-empty prediction for electric vehicles is conducted based on the participatory sensing data. Our approaches are shown to effectively improve the prediction accuracy.

Index Terms—Participatory sensing, driving energy consumption, distance-to-empty prediction, data mining

I. INTRODUCTION

Participatory sensing is an emerging paradigm of crowd-sourced data collection and knowledge discovery, which has been applied in diverse applications of pervasive and mobile computing systems [1]. The basic concept is that a group of users contribute their personal data (possibly, voluntarily) to a third-party data repository, in exchange for the useful knowledge extracted from the collective data, which is then incorporated in personalized applications of individual users.

Vehicles are becoming a vital platform for participatory sensing. First, there are extensive deployments of on-board sensors and in-vehicle information systems, equipped with network connectivity and computing power, acting as effective information collection systems. Second, the wide availability of electronic devices and smartphones carried by passengers can extend the computing and sensing abilities of vehicles. Third, there are abundant off-the-shelf and after-market automotive accessories for gathering driving data and vehicle information. Notably, participatory sensing for vehicles has been applied in several existing intelligent transportation applications (e.g., traffic status updates in Google Map and Waze).

Furthermore, advanced intelligent transportation applications can be enhanced by participatory sensing. One of the critical applications is the prediction of *distance-to-empty* (DTE) - the distance an electric vehicle (EV) or internal-combustion-engine (ICE) vehicle can reach before its energy/fuel is

exhausted. DTE is determined by a variety of factors, such as driving behavior, terrain, types of road, traffic, and vehicle specifications. The conventional approach of DTE prediction employed by car manufacturers is based on the projection of past average vehicle energy efficiency of individual drivers. Such an approach is often perceived to be inaccurate. However, if there is sufficient knowledge about the vehicle, driving behavior and the route to travel, future energy efficiency can be estimated with higher accuracy.

The availability of participatory sensing data is able to improve the accuracy of DTE prediction by exploiting the historical data from other drivers. Conceptually, one can identify the characteristics pertaining to specific driver, vehicle or environment. Then, one can harness the measurements from similar drivers, vehicles or environments to assist the prediction. In particular, there are several areas of applications:

- 1) *Vehicle Centric Applications*: Range anxiety is critical for EVs. Since there are far more ICE vehicles on the road than EVs, one can harness the data collected from ICE vehicles to improve the DTE prediction for EVs.
- 2) *Driver Centric Applications*: With diverse data collected from various drivers, one can compare the driving behavior among drivers. Hence, one can classify driving behavior and provide driving recommendations.
- 3) *Environment Centric Applications*: Eco-routing or green telematics can be provided by comparing different route options according to energy/fuel consumption metrics.

This paper presents a framework of participatory sensing by an integrated platform with appropriate knowledge discovery and incorporation mechanisms for personalized vehicle/driver/environment centric applications (depicted in Fig. 1).

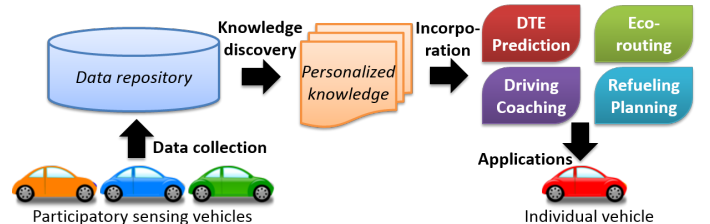


Fig. 1: An integrated platform of participatory sensing for personalized applications.

While participatory sensing can provide various driving and vehicle data, there are considerable challenges of harnessing participatory sensing. First, participatory sensing dataset is often sparse and skewed, which does not cover sufficient combinations of vehicle/driver/environment. This calls for an

effective approach to interpolate the missing data from a sparse dataset. Second, the dimensionality of dataset may be large due to different combinations of various drivers, vehicles and environments. To enable efficient data analytics, an efficient method is desirable to cope with high dimensional data.

This paper explores several approaches of utilizing participatory sensing data for personalized applications:

- 1) *Comparison with the Average*: One can obtain the average data values (e.g., average speed, stopping duration) from a large dataset for a specific environment. Then the deviation of individual drivers is compensated from the average data values in personalized applications.
- 2) *Collaborative Filtering*: A domain-free data mining technique analyzes the relationships and interdependencies within a dataset to identify a smaller set of latent factors that can characterize the observed data. Based on latent factors, one can interpolate the missing data in the dataset. In particular, matrix factorization is a critical solution to realize collaborative filtering.
- 3) *Similarity Matching*: Using a known model of driving energy consumption, one can compare the participatory sensing data and find the most similar instances from the available data to estimate the required values.

The notion of average data values has been utilized in previous papers [2], [3], [4], which has a disadvantage of requiring a large dataset. Collaborative filtering [5], [6] is a general technique in data mining without leveraging the detailed knowledge of underlying model. However, the presence of specific knowledge in driving energy consumption can potentially improve the accuracy and effectiveness. This paper presents an effective approach based on similarity matching among drivers. A blackbox approach is utilized that requires minimal knowledge of vehicles and draws on the idea of collaborative filtering. The driver/vehicle/environment dependent factors can be identified from participatory sensing data to enable personalized applications.

Various approaches for personalized prediction of driving energy consumption are compared. In particular, a case study of distance-to-empty prediction for EVs is conducted based on the participatory sensing data. Our approaches are shown to effectively improve the prediction accuracy.

Outline of Paper: The related work is first presented in Sec. II. The methodology and approaches of personalized prediction are presented in Secs. III-IV. Empirical evaluation on our approaches is given in Sec. V. The case study that utilizes our results is discussed in Sec. VI.

II. RELATED WORK

A. Driving Energy Consumption Models

Modeling driving energy consumption has been the subject of a number of research papers. One popular method is the model-based approach, which is based on vehicle dynamics to model the consumption behavior of ICE vehicles [7] and EVs [8]. Predicting energy consumption can use a blackbox approach. For example, a statistical approach using regression model to estimate the energy consumption of ICE vehicles is presented in [9]. The energy consumption rate of different

vehicles and roads can be also clustered to characterize the energy consumption of general vehicle and road types [4].

B. Data Collection of Driving Energy Consumption

The accuracy of energy consumption prediction can be enhanced by collecting more information. Two crucial factors of energy consumption prediction are the future speed profiles and future environmental factors (e.g., temperature, wind speed or route grade), which may be highly dynamic and difficult to predict. One method to estimate the future speed profiles is to utilize Markov chain [10]. Also, one can deploy sensor networks, by which stationary measurements at specific locations, such as traffic, average speed, speed limit and route grade can be measured. There are a number of papers focusing on utilizing such information [11], [3], [4]. However, the traffic data in these papers is usually static, which may have a large deviation in dynamic traffic. A study that integrates the real-time traffic sensor data to predict the energy consumption and emission of ICE vehicles is presented in [12]. One can also obtain the estimated information from social networks and participatory sensing. Participatory sensing can provide mobile measurements and good geographic penetrations [13]. For example, a study shows that the estimation of stochastic effects which impact the travel velocity and acceleration profiles can be crowd-sourced to identify traffic congestion [14]. The idea of participatory sensing has been utilized for the energy consumption prediction of ICE vehicles in [15] which shares the driving profiles among drivers. Early efforts about using social networks to identify areas of fuel reduction were reported in [16], which allow drivers to share their profiles. Our previous work employs participatory sensing for DTE prediction [17], [18], which is extended significantly in this paper.

C. Applications of Driving Energy Consumption

The integration of energy consumption prediction and data collection enables many applications. One application is the estimation of DTE, based on the prediction of the vehicle energy efficiency (i.e., energy intensity), which can be found in production vehicles [19]. DTE can be estimated by measuring the mean energy consumption over short and long distances [20]. To account for the deviation between the historical and future energy intensity, a regression model can be used to predict the future energy intensity given future route information [21]. Route features from sensor data can be clustered to identify the driving pattern for EV range estimation [22]. Another application is eco-routing. A system based on road characteristics and current prevailing traffic conditions is implemented in [23]. It can provide the users a more economic and safer route with reasonable speed instead of only driving at a lower speed. Another research utilizing average participatory sensing data and static traffic information for eco-routing system of ICE vehicles is presented in [2]. Route-type based energy consumption prediction can be implemented using OpenStreetMap (OSM) data. There are some studies using the OSM data to predict the EV range [24], [25]. A cloud based prediction system considers the deviation between the mean energy consumption and that of different condition

(e.g., traffic congestion or driving behavior) in [26]. The route-type based energy consumption model requires a complete map database including speed limit, route type and traffic information, which may not be available everywhere.

This work differentiates from the previous work in several aspects: (1) We compare various personalized prediction approaches of driving energy consumption. (2) We present a novel blackbox approach that extracts driver/vehicle/environment-dependent factors. (3) We conduct a case study of DTE prediction for EVs by different approaches.

III. METHODOLOGY AND BACKGROUND

This section presents the relevant methodology and background related to driving energy consumption models.

A. Areas of Factors

While there are many factors to determine driving energy consumption, they can be classified by three broad areas:

- **Driver:** The driver who controls the vehicle has a direct impact on the vehicle movement. Different drivers exhibit different preferences for stop/start and acceleration, aggression in various scenarios, propensity for hypermiling, etc. Psychological and behavioral traits of drivers also affect driving energy efficiency.
- **Vehicle:** Different types of vehicles consume energy differently. ICE vehicles are characterized by the engine types and gear shifts, whereas hybrid and EVs are affected by battery performance and regenerative braking. The sizes and weights of vehicles often determine the efficiency of kinetic energy conversion, so SUVs and trucks are usually less energy-efficient than sedan and compact vehicles.
- **Environment:** The environmental factors include traffic and roads. Traffic for a road segment depends on a plethora of factors, including time-of-day, day-of-year, special events, which may follow a certain pattern. The types of roads also affect drivers' behavior differently, which can be divided into three main categories: small public or private roads with urban traffic, lower capacity "urban" highways, and higher capacity freeways. Other environmental factors, such as road grades and weather types, can also be considered.

The historical data of vehicle speed profiles can be identified by a combination of (driver, vehicle, environment), referred as a *data point*. This paper aims to predict the energy consumption for each data point. Through participatory sensing, a dataset of measured energy consumption for a relatively small number of data points are collected. This paper addresses the challenge of data interpolation with good accuracy.

B. Types of Models

There are two main types of energy consumption models:

- **White-box Model:** A straightforward approach is to employ a white-box microscopic behavior model of each vehicle that comprehensively characterizes the engine performance, vehicle mechanics, battery systems, etc. To incorporate traffic information, one can rely on a macroscopic traffic database collected from a network

of loop sensors along specific road segments. However, such a white-box vehicle model requires a large amount of data for calibration and detailed knowledge specific to a particular vehicle. Also, the availability and access of accurate traffic information is often limited to certain authorized parties only.

- **Blackbox Model:** A blackbox approach is more desirable that requires minimal knowledge of vehicle model with only a small set of measurable variables and parameters. The variables and parameters depend on the combinations of (driver, vehicle, environment). In the subsequent sections, the variables and parameters obtained in the blackbox model will be utilized for collaborative filtering and similarity matching.

C. Energy Consumption Model

This section describes a linear blackbox model of driving energy consumption that has been used extensively in the literature [2], [4], [9], [17], [27]. Denote a driver by D , a vehicle model by V , and a particular environment (e.g., a segment of route and time-of-day) by R . Each energy consumption is represented by a numerical value $E_{D,V,R}$, indexed by the tuple (D, V, R) . All the entries of energy consumption values form a 3-dimensional tensor, denoted by $[E_{D,V,R}]$.

While there are sophisticated approaches of estimating the moving vehicle energy consumption by white-box microscopic behavior models [8], [7], [10], [22], [28], [29], these models are rather difficult to implement. Many parameters are required, for example, engine efficiency, transmission efficiency, regenerative braking efficiency, etc. However, in practice, these parameters are hard to obtain. Therefore, this paper utilizes a blackbox approach without the detailed knowledge of vehicle mechanics. This approach maximizes the applicability for a wide range of scenarios arising from participatory sensing.

The total energy consumption E of driver D with vehicle model V in a particular environment R is given by:

$$E_{D,V,R} = E_{D,V,R}^{mv} + E_{D,V,R}^{id} \quad (1)$$

where $E_{D,V,R}^{mv}$ is the moving vehicle energy consumption and $E_{D,V,R}^{id}$ is the idle vehicle energy consumption. Simple blackbox models are next presented to characterize $E_{D,V,R}^{mv}$ and $E_{D,V,R}^{id}$.

1) Moving Vehicle Energy Consumption:

With respect to a particular combination of (D, V, R) , the moving vehicle energy consumption E^{mv} has unit in liter or kWh. Next, the subscript D,V,R is dropped for brevity.

In this paper, E^{mv} (denoted by \hat{E}^{mv}) is estimated by a linear equation of several measurable variables from vehicles¹:

$$\hat{E}^{mv} = \begin{bmatrix} \alpha_{v,1} \\ \alpha_{v,2} \\ \vdots \\ \alpha_{v,r} \end{bmatrix}^T \begin{bmatrix} v \\ v^2 \\ \vdots \\ v^r \end{bmatrix} + \begin{bmatrix} \vec{\alpha}_{d,1} \\ \vec{\alpha}_{d,2} \\ \vdots \\ \vec{\alpha}_{d,k} \end{bmatrix}^T \begin{bmatrix} \vec{d} \\ \vec{d}^2 \\ \vdots \\ \vec{d}^k \end{bmatrix} + \begin{bmatrix} \vec{\alpha}_{a,1} \\ \vec{\alpha}_{a,2} \\ \vdots \\ \vec{\alpha}_{a,m} \end{bmatrix}^T \begin{bmatrix} \vec{a} \\ \vec{a}^2 \\ \vdots \\ \vec{a}^m \end{bmatrix} + \begin{bmatrix} \alpha_g \\ \alpha_\ell \\ c \end{bmatrix}^T \begin{bmatrix} g \\ \ell \\ 1 \end{bmatrix} \quad (2)$$

¹Some of the variables are selected based on [30], which analyzed more than 20 thousand data points from 45 drivers to identify the most significant factors of fuel consumption and emission.

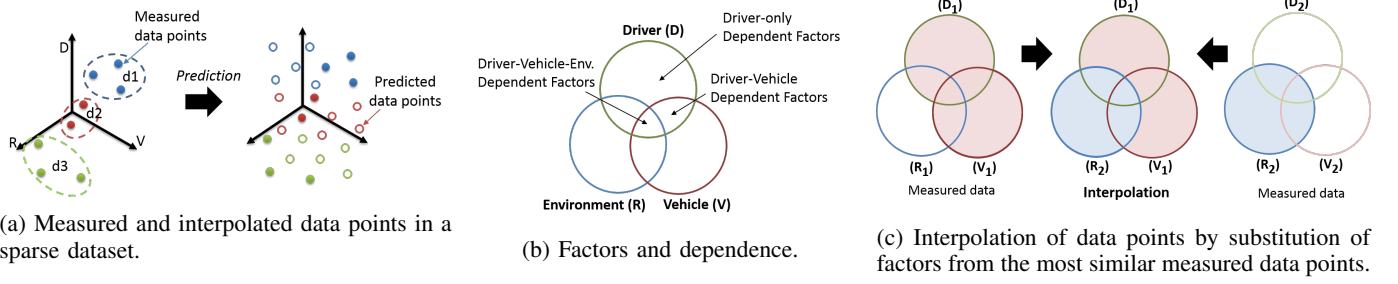


Fig. 2: Illustrations for interpolation of missing data points.

where

- v is the continuous average speed (i.e., the average speed without idling). The higher powers² of v like v^2, \dots, v^r are also considered.
- $\vec{d} = (\tau_d, \mu_d, \sigma_d)$ is the deceleration tuple:
 - τ_d is the total duration of deceleration.
 - μ_d is the mean deceleration (i.e., the sum of deceleration values divided by the deceleration duration).
 - σ_d is the standard deviation of deceleration.

Denote the higher powers of components in the deceleration tuple by $\vec{d}^k = (\tau_d^k, \mu_d^k, \sigma_d^k)$.

- \vec{a} is the acceleration tuple (similar to \vec{d}).
- g is the mean absolute value of gyroscope along the moving direction.
- ℓ is the auxiliary load of idling, which is the baseline measurement when the vehicle is not moving.
- c is a normalization constant.
- $\alpha_v \triangleq (\alpha_{v,1}, \dots, \alpha_{v,r})$, $\alpha_d \triangleq (\alpha_{d,1}, \dots, \alpha_{d,k})$, $\alpha_a \triangleq (\alpha_{a,1}, \dots, \alpha_{a,k})$, α_g, α_ℓ are the corresponding coefficients.

2) Idle Vehicle Energy Consumption:

Similarly, a blackbox approach is used to estimate the idle vehicle energy consumption. The subscript D, V, R is dropped for brevity. With respect to a particular combination of (D, V, R) , the idle vehicle energy consumption E^{id} (denoted by \hat{E}^{id}) is estimated by a linear equation:

$$\hat{E}^{\text{id}} = \beta_1 \mu \ell + \beta_2 \omega \quad (3)$$

where

- μ is the total idle duration.
- ℓ is the auxiliary load of idling.
- ω is the outdoor temperature.
- β_1, β_2 are the coefficients.

The parameters v, ℓ can be obtained from standard OBD data inquiry from vehicles, whereas \vec{d}, \vec{a}, μ can be computed from speed profiles, g can be obtained from smartphones, and ω can be obtained from online weather data.

D. Estimation of Coefficients

The coefficients $(\alpha_v, \vec{\alpha}_d, \vec{\alpha}_a, \alpha_g, \alpha_\ell, c, \beta_1, \beta_2)$ in Eqns. (2)-(3) can be estimated by the standard regression method, if sufficient measured data $(v, \vec{d}, \vec{a}, g, \ell, \mu, \omega)$ and the respective energy consumption data $(\hat{E}^{\text{mv}}, \hat{E}^{\text{id}})$ are provided. Assume

²Sec. V will empirically determine the proper powers of parameters.

that each driver-vehicle pair (D, V) has collected sufficient historical personal driving data, and hence, the coefficients can be estimated for the respective environment R . One notable advantage of regression method is that it is less susceptible to random noise, which can arise from various sources (e.g., due to time synchronization in data sampling, mechanic damping, inaccurate measurements).

IV. INTERPOLATING PARTICIPATORY SENSING DATA

Given a dataset of driving and vehicle data collected by participatory sensing, a data point can be visualized as a point in a 3-dimensional Euclidean space, indexed by (D, V, R) . The participatory sensing dataset is usually sparse, consisting of a skewed and clustered distribution of data points. In order to predict the driving energy consumption for the data points that are not collected from participatory sensing, we seek to interpolate the missing data points to cover the space of dataset. An illustration is depicted in Fig. 2a.

Next, three major data interpolation approaches by similarity matching, matrix factorization, and comparison with the average are presented.

A. Similarity Matching

The three areas of factors (i.e., driver, vehicle, and environment) that determine the driving energy consumption are not necessarily exclusive. There are factors that can belong to multiple aspects. For example, the speed of a vehicle depends on both driver and environment. Abstractly, the factors can be visualized by a Venn diagram (see Fig. 2b). Similarity matching is related to neighborhood-based collaborative filtering.

After characterizing the factors and their dependence, the interpolation of missing data points can be attained by suitable substitution of factors from the most similar measured data points. For example, see Fig. 2c for an illustration. After obtaining the measured data for (D_1, V_1, R_1) and (D_2, V_2, R_2) , we aim to estimate the energy consumption for (D_1, V_1, R_2) . If D_1 is similar to D_2 and V_1 is similar to V_2 , then one can replace the factors that depend on R_1 in (D_1, V_1, R_1) by those depend on R_2 in (D_2, V_2, R_2) .

The energy consumption model in Eqns. (2)-(3) provides a convenient way to extract the factors of driver, vehicle, and environment dependence. In Table I, the dependence of each parameter is heuristically assigned based on the major observable impacts from the driver, vehicle or environment.

For the coefficients, it is assumed that their dependence is complementary to that of the respective parameters. For

	Driver-dependent	Vehicle-dependent	Environment-dependent
$v, \vec{d}, \vec{a}, g, \ell$	✓		✓
μ, ω			✓
$\alpha_v, \vec{\alpha}_d, \vec{\alpha}_a, \alpha_g$		✓	
α_ℓ, c		✓	
β_1, β_2	✓	✓	

TABLE I: Dependence of parameters and coefficients.

example, the average speed v is more likely affected by the driver and environment, while to a less extent by the type of vehicle. Hence, coefficient α_v is considered to be vehicle-dependent, such that the product $\alpha_v v$ will be specific to a particular tuple (D, V, R) . The dependence will be empirically validated in Sec. V.

The interpolation of missing data points can be attained by the substitution of parameters and coefficients in the driving energy consumption model. For example, consider Fig. 2c. Let the parameters and coefficients for $(D1, V1, R1)$ be $(\alpha_v, \vec{\alpha}_d, \vec{\alpha}_a, \alpha_g, \alpha_\ell, c, \beta_1, \beta_2)$ and $(v, \vec{d}, \vec{a}, g, \ell, \mu, \omega)$, and those for $(D2, V2, R2)$ be $(\alpha'_v, \vec{\alpha}'_d, \vec{\alpha}'_a, \alpha'_g, \alpha'_\ell, c', \beta'_1, \beta'_2)$ and $(v', \vec{d}', \vec{a}', g', \ell', \mu', \omega')$. To estimate the energy consumption for $(D1, V1, R2)$, $(\alpha_v, \vec{\alpha}_d, \vec{\alpha}_a, \alpha_g, \alpha_\ell, c, \beta_1, \beta_2)$ and $(v', \vec{d}', \vec{a}', g', \ell', \mu', \omega')$ are used in the driving energy consumption model.

To determine the similarity among drivers and vehicles, two approaches of similarity matching by speed profile matching and driving habit matching are presented next.

1) Speed Profile Matching:

One can characterize the similarity between a pair (D, V) and (D', V') under the same environment R by comparing the respective speed profiles (i.e., the plots of speed against traveled distance). Since speed profiles are time series, *dynamic time warping* (DTW) [31] can be used as a metric for determining the similarity, and identifying the corresponding similar regions between two time series, which has been applied in many applications (e.g., speech recognition).

The basic idea of DTW is to determine an optimal alignment between two time series. Consider two time series $X = (x[t])_{t=1}^{n_X}$ and $Y = (y[t])_{t=1}^{n_Y}$ of lengths n_X and n_Y respectively. A *warp path* is defined as $W = (w[k])_{k=1}^{n_W}$, where the k -th element is $w_k = (i, j)$, such that i is an index from time series $x[i]$ and j is an index from time series $y[j]$. n_W is the length of the warp path W , such that $\max(n_X, n_Y) \leq n_W < n_X + n_Y$. The warp path W is subject to the following constraints:

- 1) $w[1] = (1, 1)$ and $w[n_W] = (n_X, n_Y)$;
- 2) if $w[k] = (i, j)$ and $w[k+1] = (i', j')$, then $i \leq i' \leq i+1$ and $j \leq j' \leq j+1$.

The warp path of minimum distance $\text{dist}(W^*)$ is defined by:

$$\text{dist}(W^*) = \min_w \sum_{k=1}^{n_W} d(w[k]) \quad (4)$$

where each $d(w[k]) = |x[i] - y[j]|$ is the distance of the coordinates (i, j) of the k -th element in W . A simple approach to determine an optimal warp path between two time series is

using dynamic programming. But there are other more efficient algorithms with linear running time [31].

Suppose each trip is divided into a sequence of segments (R^i) . Let $v_{D,V,R^i}[t]$ be the time series of speed profile for tuple (D, V, R^i) . For each pair of (D, V, R^i) and (D', V', R^i) , define

$$\chi_{(D,V),(D',V')}^{R^i} \triangleq \text{dist}(W^*) \quad (5)$$

where W^* is the minimum-distance warp path between the time series $v_{D,V,R^i}[t]$ and $v_{D',V',R^i}[t]$.

Let $\mathcal{R}(D, V)$ be a set of segments that have speed profiles measured with (D, V) . Namely, if $R^i \in \mathcal{R}(D, V)$, then the speed profile $v_{D,V,R^i}[t]$ exists in the dataset. Define a similarity metric between each pair of (D, V) and (D', V') by the average minimum warp path distance over all segments:

$$\bar{\chi}_{(D,V),(D',V')} \triangleq \frac{\sum_{R^i \in \mathcal{R}(D,V) \cap \mathcal{R}(D',V')} \chi_{(D,V),(D',V')}^{R^i}}{|\mathcal{R}(D,V) \cap \mathcal{R}(D',V')|} \quad (6)$$

Note that $\bar{\chi}_{(D,V),(D',V')} = \infty$, if $\mathcal{R}(D,V) \cap \mathcal{R}(D',V') = \emptyset$.

For example, the speed profiles of three driver-vehicle pairs (D_1, V_1) , (D_2, V_2) , (D_3, V_3) on the same trip R^1 are plotted in Fig. 3. Smaller minimum warp path distance is observed to have closer similarity in the speed profiles - (D_1, V_1) is more similar to (D_2, V_2) than (D_3, V_3) .

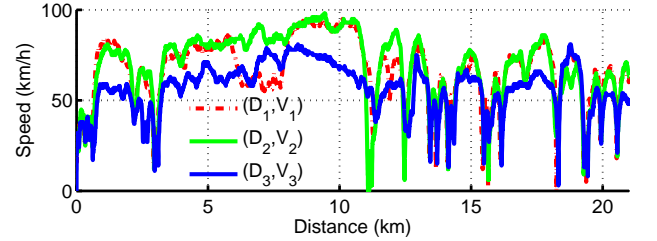


Fig. 3: Speed profiles of three drivers on the same trip. $\chi_{(D_1,V_1),(D_2,V_2)}^{R^1} = 1.1385$ and $\chi_{(D_1,V_1),(D_3,V_3)}^{R^1} = 1.3883$.

This paper uses $\bar{\chi}_{(D,V),(D',V')}$ to characterize the similarity between each pair of (D, V) and (D', V') . The tuple (D', V') with the smallest value $\bar{\chi}_{(D,V),(D',V')}$ is identified for estimating energy consumption of (D, V) . For finding multiple similar data points, k -nearest neighbors (k -NN) clustering is employed to find the k most similar speed profiles with (D, V) .

2) Driving Habit Matching:

Speed profiles are not always available for the same environment. An alternative is to rely on the available data collected from other environments. An important factor for driving energy consumption is the acceleration/deceleration [30]. The accelerating behavior of the drivers is related to driving energy efficiency. On the other hand, aggressive decelerations, usually inducing rear-end collisions, is related to driving awareness.

For example, the mean acceleration and deceleration (μ_d, μ_a) against the continuous average speed of each segment from the data of a driver are plotted in Fig. 4. It is observed that the decelerations/accelerations tend to be higher at a low speed (possibly, due to stop-and-go behavior), whereas lower decelerations/accelerations can be found at a high speed. Low

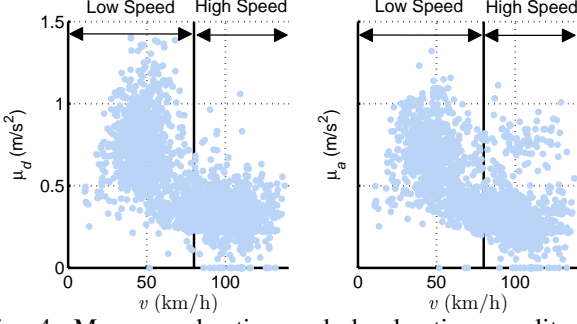


Fig. 4: Mean acceleration and deceleration amplitudes vs. continuous average speed of each segment of a driver.

accelerations are usually due to cruise control or mindful drivers, while high accelerations are due to aggressive driving.

Therefore, we are motivated to use the average acceleration and deceleration as a metric to characterize driving habits. However, we may not have collected sufficient measurements for every vehicle speed. Hence, normalize the distribution of data to obtain a better estimation of the average. First, divide the range of vehicle speed into a sequence of intervals with width Δv (i.e., $[v, v + \Delta v]$). This paper considers $\Delta v = 10\text{km/h}$. For each interval $[v, v + \Delta v]$, let $\gamma_a^v(D, V)$ be the mean value of the acceleration measurements within $[v, v + \Delta v]$. Define the *estimated average acceleration* by the average of the mean values in all intervals by $\bar{\gamma}_a(D, V)$.

Because a difference is observed in high-speed and low-speed driving habits, define different estimated average accelerations for the intervals above or below a threshold v_{th} :

- 1) Low-speed estimated average acceleration $\bar{\gamma}_a^{low}(D, V)$.
- 2) High-speed estimated average acceleration $\bar{\gamma}_a^{high}(D, V)$.

Similarly, define $\bar{\gamma}_d^{low}(D, V)$ and $\bar{\gamma}_d^{high}(D, V)$ for deceleration. Fig. 5 depicts an illustration of $\gamma_d^v(D, V)$, $\bar{\gamma}_d^{low}(D, V)$ and $\bar{\gamma}_d^{high}(D, V)$.

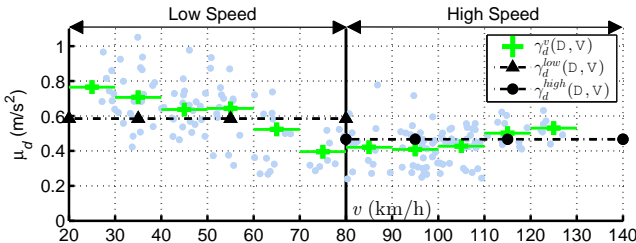


Fig. 5: An illustration of $\bar{\gamma}_d(D, V)$ and $\gamma_d^v(D, V)$.

Heuristically, set $v_{th} = 80\text{km/h}$, because this speed limit usually sets the difference between highways and suburban roads. The average deceleration/acceleration tuple $(\bar{\gamma}_a^{low}(D, V), \bar{\gamma}_a^{high}(D, V), \bar{\gamma}_d^{low}(D, V), \bar{\gamma}_d^{high}(D, V))$ can capture the driving habit of each (D, V) . The average deceleration/acceleration tuple is used to compare the similarity between each pair (D, V) and (D', V') .

B. Matrix Factorization

The similarity matching approaches are based on domain-specific knowledge. Collaborative filtering is a domain-free approach, relying on the identification of abstract latent factors. Matrix factorization is a popular approach of constructing

latent factors, which has been implemented in recommendation system [6] and other large-scale problems [5].

Consider an example of sparse matrix R of n pairs of (D, V) and m road segments R , as shown in Table II, in which each entry represents a measurement (e.g., v , \bar{d} or \bar{a}). Note that some data points may be missing in R .

$D, V \backslash R$	1	2	3	4	...	m
1	67	74	-	32	...	50
2	54	-	83	44	...	65
\vdots						
n	-	66	58	-	...	88

TABLE II: An example of sparse matrix R of vehicle speed v .

The basic idea of matrix factorization is to find two low-rank $(n \times k$ and $m \times k)$ matrices, P and Q , such that PQ^T can approximate R . Namely,

$$R \approx PQ^T = \hat{R} \quad (7)$$

P and Q can be regarded as mappings to reduce the m, n -dimensional space of the original dataset to a k -dimensional space of latent factors, where $k \ll \min(m, n)$. Denote the entry at the i -th column and the j -th row of R be r_{ij} .

The objective of matrix factorization is find P, Q such that

$$\min_{P, Q} \sum_{i, j} (r_{ij} - p_i q_j^T)^2 + \lambda_P \|p_i\|^2 + \lambda_Q \|q_j\|^2 \quad (8)$$

where p_i is the i -th row vector of P , and q_j is the j -th column vector of Q . Since factorization may cause over-fitting, λ_P and λ_Q are used to regularize the fitting.

There are two popular approaches to compute P, Q in Eqn. (8): stochastic gradient descent [5] and alternating least squares [6]. In this paper utilizes stochastic gradient descent. The basic idea is to go through all r_{ij} in R . For each r_{ij} , determine the corresponding factor vectors p_i and q_j . Then, compute the approximate value by $p_i q_j^T$ and update the parameters by:

$$\begin{aligned} p_i &\leftarrow p_i + \epsilon(e_{ij}q_j - \lambda_P p_i) \\ q_j &\leftarrow q_j + \epsilon(e_{ij}p_i - \lambda_Q q_j) \end{aligned} \quad (9)$$

where $e_{ij} = r_{ij} - p_i q_j^T$ represents the difference between approximate value and actual value and ϵ is the learning rate. Once P, Q are determined, the estimation of a missing data \hat{r}_{ij} can be estimated by $\hat{r}_{ij} = p_i q_j^T$. All measurements (e.g., v , \bar{d} or \bar{a}) can be substituted and estimated using matrix factorization. The estimated values can be utilized in the driving energy consumption prediction.

C. Comparison with the Average

A basic approach for estimating driving energy consumption is based on the global average data values (e.g., average speed) from participatory sensing data. However, each driver may deviate considerably from the average data values. To compensate for the deviations, incorporating a personalized adjustment can improve the prediction accuracy.

Let $f_{D, V, R}$ be a personal data value for tuple (D, V) in environment R , and the average data value be \bar{f}_R . An adjustment

function $\mathcal{D}_{D,V}^f(\cdot)$ is used to convert the average data value to the personal data value by:

$$f_{D,V,R} = \mathcal{D}_{D,V}^f(\bar{f}_R) \quad (10)$$

In this paper, a simple adjustment function is considered by regression model:

$$\mathcal{D}_{D,V}^f(\bar{f}_R) = \eta_1 \bar{f}_R^2 + \eta_2 \bar{f}_R + \eta_3 \quad (11)$$

V. EMPIRICAL EVALUATION

Empirical evaluation on the energy consumption model and its properties are discussed in this section.

A. Setup

The data from 5 drivers and 7 vehicles is collected. The information of vehicles is given in Table III. Some drivers drove multiple vehicles, which gives totally 10 tuples of (D,V). Since the scenario of participatory sensing is considered, it suffices to consider a relatively small dataset.

Vehicle	Maker	Model	Year	Type	Displacement
V ₁	Nissan	LEAF	2014	EV	NA
V ₂	Ford	Fiesta	2013	ICE	1.4
V ₃	Toyota	Yaris	2013	ICE	1.5
V ₄	Hyundai	Veloster	2014	ICE	1.6
V ₅	Ford	Fusion	2012	ICE	2.5
V ₆	BMW	650i	2014	ICE	5.0
V ₇	Ford	F150	2014	ICE	5.0

TABLE III: The vehicles in the experiments.

Totally 3000 km of data is collected. Fig. 6 depicts the distance of collected data for all driver-vehicle pairs. The data is then are segmented to 1-km segments.

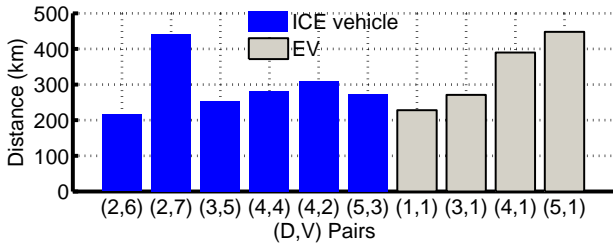


Fig. 6: Collected data of all driver-vehicle pairs.

For ICE vehicles, we collected data through ELM327 devices connected to vehicles' onboard diagnostic (OBD) ports and paired with a smartphone. The collected OBD data include mass air-flow, manifold absolute pressure, intake air temperature and engine RPM. Geo-location data, accelerometer and gyroscope measurements from the smartphone are also collected. For EVs (i.e., Nissan LEAF), resolution state-of-charge (SOC) and vehicle speed data is collected.

B. Estimation Errors of Energy Consumption Model

The ground truth energy consumption data (i.e., $E_{D,V,R}$, $E_{D,V,R}^{mv}$, $E_{D,V,R}^{id}$) can be obtained from OBD data. From the OBD data of ICE vehicles, the fuel rate can be estimated based on mass air flow and fuel/air ratio [32]. From the OBD data of EVs, the energy consumption is estimated by SOC and the

battery capacity. Readers can refer to [33] for the details of extraction OBD data from EVs.

Two metrics of error are utilized to evaluate the energy consumption predictions in this study. The first metric of error is the per-segment error for each segment of road R^i :

$$\varepsilon^i = \frac{(E_{D,V,R^i}^{mv} + E_{D,V,R^i}^{id}) - (\hat{E}_{D,V,R^i}^{mv} + \hat{E}_{D,V,R^i}^{id})}{E_{D,V,R^i}^{mv} + E_{D,V,R^i}^{id}} \quad (12)$$

which is used to evaluate the accuracy of the energy consumption model (Eqns. (2)-(3)). The second metric of error is the accumulative error:

$$\varepsilon^{acc} = \frac{|E_{D,V,R} - \hat{E}_{D,V,R}|}{E_{D,V,R}} \quad (13)$$

which is used to evaluate the energy prediction accuracy over a trip composed of many segments. Fig. 7 shows the speed profile and energy consumption of an ICE vehicle and an EV data. The idling energy consumption ($E_{D,V,R}^{id}$) is identified from speed profile, and the moving energy consumption ($E_{D,V,R}^{mv}$) is obtained by Eqn. (1).

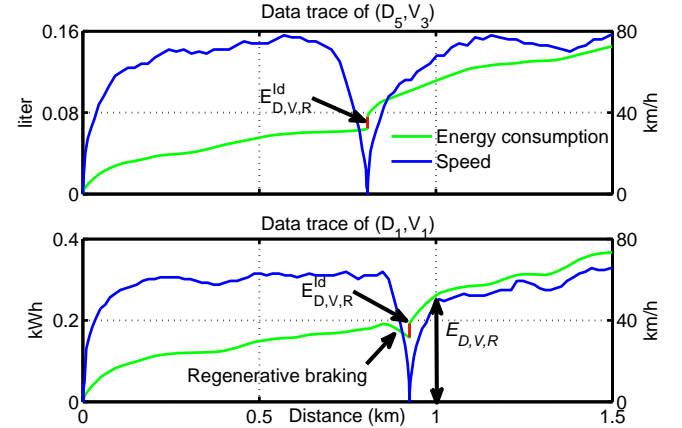


Fig. 7: Energy consumption of different driver-vehicle pairs.

C. Fitness of Energy Consumption Model

In this section, the proper powers of v, \vec{d}, \vec{a} in Eqn. (2) for model fitting are evaluated. The Akaike Information Criterion (AIC) [34] is utilized to determine the proper values of (r, k, m) . AIC estimates the quality of each model and balances the trade-off between the goodness of model fitting and the complexity of model. The AIC value of a model can be computed using the estimated residual in least square method. Consider the energy consumption model using (r, k, m) order of powers in Eqn. (2), the AIC value is expressed by:

$$AIC_{(r,k,m)} = n \log \frac{\sum \varepsilon_i^2}{n} + 2K \quad (14)$$

where ε is the per-segment error (see Eqn. (12)), n is the number of segment and K is the total number of estimated regression coefficients (e.g., $r + k + m + 3$). According to AIC test criterion, the smaller value makes the better model. The AIC value is normalized as $AIC_{(r,k,m)}^N$, with respect to the smallest AIC value in all driver-vehicle pairs ($\min(AIC)$):

$$AIC_{(r,k,m)}^N = \frac{AIC_{(r,k,m)}}{\min(AIC)} - 1 \quad (15)$$

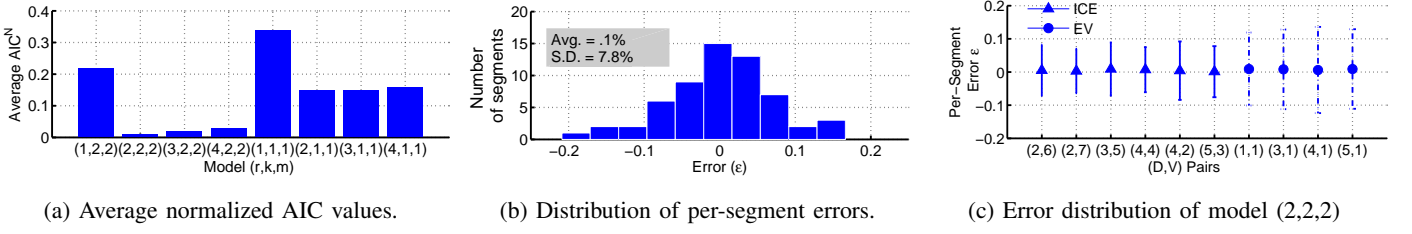


Fig. 8: Experimental data and evaluation.

The mean normalized AIC values averaged over all driver-vehicle pairs for a particular (r, k, m) are plotted in Fig. 8a, which shows that $(2, 2, 2)$ attains the minimum, and hence, is the best setting of powers of v, \vec{d}, \vec{a} in Eqn. (2).

D. Evaluation of Energy Consumption Model

The per-segment errors of $(2, 2, 2)$ model for all (D, V) pairs are validated in this section. 80% of collected data (called in-sample data) are randomly selected to train the regression model. The rest of data (called out-sample data) are utilized to validate the accuracy of the model. The per-segment error distribution of out-sample data for driver-pair pair $(4, 4)$ is shown in Fig. 8b. The total number of segment is 58 (58km). The mean error is 0.1%. The standard deviation is about 7.8% and the distribution approaches a normal distribution. This shows that our energy consumption model is relatively accurate. The model validation results of all pairs are displayed in Fig. 8c. Slightly higher standard deviation of per-segment error is observed for EVs, because of a lower sample rate.

Since we are interested in the energy consumption of the overall trip, the accumulative error is more relevant. Fig. 9 shows the accumulative error against traveled distance over multiple rounds in the same route. It is observed that although the standard deviation is up to 5%, the accumulative error is much smaller. This is due to the fact that the positive and negative deviations can offset each other over a longer distance. Therefore, the accumulative error has a lower value after a longer distance. The root mean square accumulative error $\text{RMSE}(\epsilon^{\text{acc}})$, which measures the performance over the traveled distance, is also examined. In the case study (Sec. VI), $\text{RMSE}(\epsilon^{\text{acc}})$ is used to evaluate the accuracy of energy consumption prediction for a designated trip.

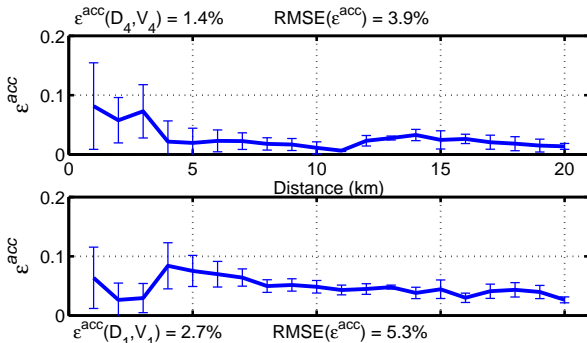


Fig. 9: Accumulative error against traveled distance.

E. Dependence of Coefficients

To properly assign the dependence of coefficients in the energy consumption model in Eqns. (2)-(3), the distribution of coefficients between all driver-vehicle pairs is examined to identify the dependence empirically. To compare between different energy resource, the suggested conversion from the US Environmental Protection Agency (US EPA) is utilized to convert the kwh to gasoline fuel, in which 33.7 kilowatt hours of electricity is equivalent to one gallon of gasoline [35].

To validate the dependence, first randomly draw a portion (80%) of training data to train the model for each driver-vehicle pair, and repeat the procedure 100 times to create 100 sets of coefficients for each pair. As an example, the distributions of coefficients $\alpha_{v,2}$ and c to the same driver or the same vehicle are plotted in Fig. 10. It is observed that the distributions for of coefficients $\alpha_{v,2}$ and c of the same driver in different vehicles tends to be independent from another vehicle. In addition, the rightmost figures show the distributions of different drivers in the same vehicle are highly overlapping, which means the coefficients are less affected by drivers. Therefore, the coefficients $\alpha_{v,2}$ and c are assigned to be vehicle dependent. The dependence of other parameters and coefficients are also validated in Table I.

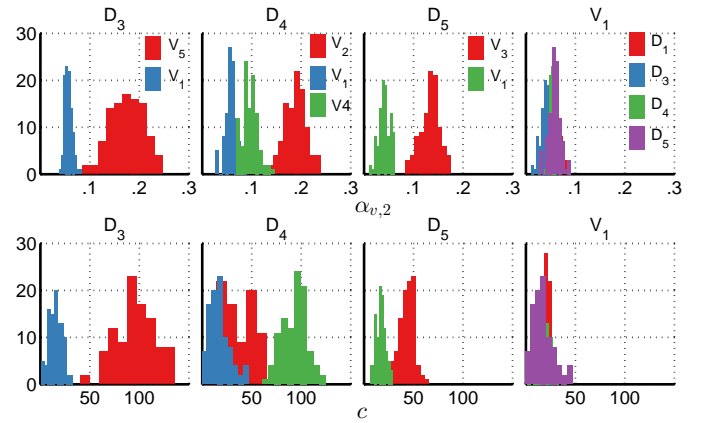


Fig. 10: Distributions of coefficients $\alpha_{v,2}$ and c .

F. Driving Habits

This section compares the driving habits characterized by low-speed and high-speed average acceleration/deceleration tuple. The average accelerations/decelerations of several drivers are plotted in Fig. 11, which are aggregated over multiple trips. Positive correlations between average acceleration and average deceleration is observed. Drivers who

accelerate more tend to decelerate more. As a result, one can classify the driving habits by awareness and efficiency according to different regions in low-speed and high-speed average accelerations/decelerations plots, relative to the mean values among drivers.

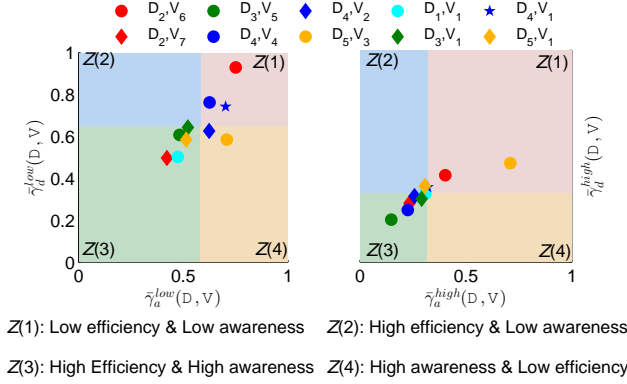


Fig. 11: Regions of driving habits characterized by average accelerations/decelerations.

VI. CASE STUDIES

This section presents the case studies of various personalized prediction approaches. All driver-vehicle pairs are required to drive in a designated route for evaluation. The ground truth energy consumption data is also collected. Fig. 12 shows the energy consumption and speed profiles for two driver-vehicle pairs. The designated route comprises of suburban (0 to 20 km) and stop-and-go (20 to 31 km) parts. 3 rounds of driving are repeated to obtain training data and reference data.

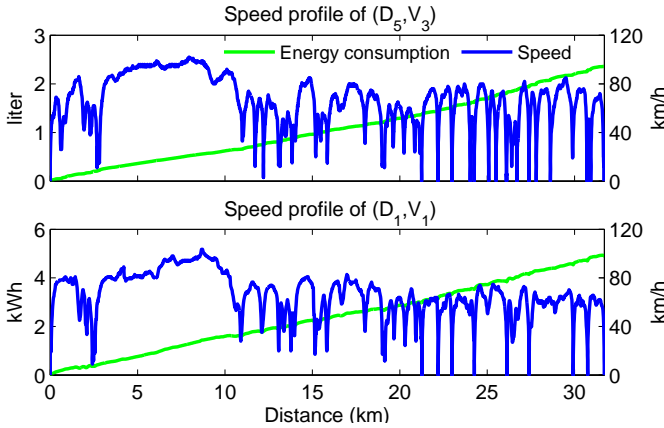


Fig. 12: Energy consumption and speed profiles of the designated route in the case study.

A. Personalized Driving Energy Consumption Prediction

This section compares the performance of various personalized prediction approaches for driving energy consumption using the collected data for the designated route. The energy consumption model of each driver-vehicle pair is trained using the historical data collected from daily driving. Then the

energy consumption of the route can be predicted using the collected data from different driver-vehicle pairs for the route.

For example, the energy consumption model for \hat{E}^{mv} of (D_1, V_1) of a particular road segment is obtained from the historic data and is given as follows.

$$\hat{E}^{mv} =$$

$$\begin{bmatrix} -5.831 \\ 0.266 \\ 0.483 \\ 14.863 \\ 76.918 \end{bmatrix}^T \begin{bmatrix} v \\ v^2 \\ g \\ \ell \\ 1 \end{bmatrix} + \begin{bmatrix} -6.261 \\ 143.340 \\ -139.219 \\ 0.034 \\ -30.534 \\ 86.334 \end{bmatrix}^T \begin{bmatrix} \tau_d \\ \mu_d \\ \sigma_d \\ \tau_d^2 \\ \mu_d^2 \\ \sigma_d^2 \end{bmatrix} + \begin{bmatrix} 11.067 \\ 11.425 \\ -12.059 \\ -0.056 \\ -120.340 \\ 169.864 \end{bmatrix}^T \begin{bmatrix} \tau_a \\ \mu_a \\ \sigma_a \\ \tau_a^2 \\ \mu_a^2 \\ \sigma_a^2 \end{bmatrix}$$

Various personalized prediction approaches are considered as follows:

- 1) *Speed Profile Matching* (SPM): The paths among driver-vehicle pairs from historical data are matched using GPS data. The distance metric $\bar{\chi}_{(D,V),(D',V')}$ is computed using Eqn. (6) for all pairs of (D, V) and (D', V') . Besides, k -nearest neighbors (k -NN) clustering is utilized to determine k nearest pairs in distance metric $\bar{\chi}_{(D,V),(D',V')}$. The similarity matching approach based on k nearest pairs is denoted by SPM(k).
- 2) *Driving Habit Matching* (DHM): The low-speed and high-speed average deceleration/acceleration tuple $(\bar{\gamma}_a^{\text{low}}(D, V), \bar{\gamma}_a^{\text{high}}(D, V), \bar{\gamma}_d^{\text{low}}(D, V), \bar{\gamma}_d^{\text{high}}(D, V))$ are computed for every driver-vehicle pair (D, V) . The low-speed and high-speed average deceleration/acceleration tuple defines a 4-dimensional data space. The similarity matching approach based on k nearest pairs in the 4-dimensional data space is denoted by DHM(k).
- 3) *Matrix Factorization* (MF): The paths among driver-vehicle pairs from collected data using GPS data are matched before employing matrix factorization. The matrix factorization approach is denoted by MF.
- 4) *Average Data Values* (Avg.): Using only the average data values are used (e.g., average speed) for driving energy consumption prediction. Sometimes, the average speed is observed to be very close to the speed limit. The average data based approach is denoted by Avg.
- 5) *Adjusted Personal Data Values* (Adj.): The adjustment function in Eqn. (11) is used to convert the average data values to the personal data values. The adjusted personal data based approach is denoted by Adj.
- 6) *Self-Estimation* (Self Est.): Using only one's own data in energy consumption model of the same road segment in Eqns. (2)-(3) is also considered. The self-estimation approach is denoted by Self Est. Self-estimation is a benchmark, which essentially validates the accuracy of the energy consumption model without using the data of other driver-vehicle pairs. In practice, one's own data of the same road segment may not be always present, as the driver has not traveled such a route before.

Fig. 13 compares the prediction errors in terms of RMSE, against the ground truth energy consumption over all driver-vehicle pairs. The strengths and weaknesses of each prediction approach are summarized in Table IV.

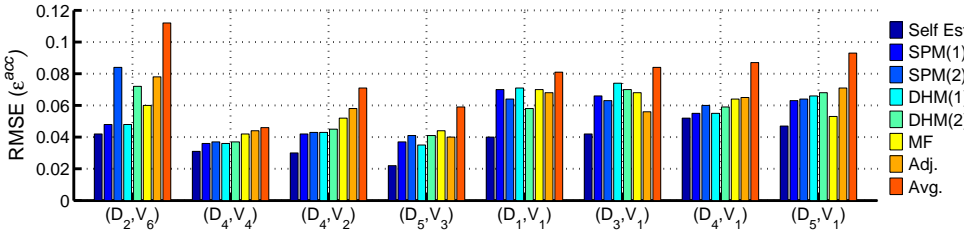


Fig. 13: Prediction error $RMSE(\varepsilon^{acc})$ over all (D, V) pairs.

Self Est. is observed to have 5% error, which is the lowest among all approaches, because one's own driving data on the same route is the most accurate source for prediction, in spite of the presence of different traffic condition. SPM is observed to have a close prediction error with Self Est. Avg. is observed to have the largest error, because of the considerable deviation from individual drivers from the average. Adj. can improve the accuracy of comparison with the average. Notably, DHM is observed to perform relatively well, even though it does not require path matching using GPS data. Therefore, driving habits are a good indicator of driving energy consumption. In summary, DHM provides good accuracy without GPS data, which has low complexity for system implementation.

B. Distance-to-Empty Prediction for EV

In this section, our approaches are applied to the application of DTE prediction for EV (i.e., Nissan LEAF) using other ICE vehicle data. For the convenience of comparison, certain routes are selected and all drivers are required to travel the same route at least 3 times for evaluations.

The data collected from Nissan LEAF includes:

- 1) State-of-charge (SOC), denoted by S , which indicates the remaining battery level.
- 2) Initial capacity of the battery, denoted by B_A .
- 3) Battery pack voltage when driving, denoted by B_V .

The remaining energy ($\Delta\mathcal{E}^t$) in battery at time t is given by:

$$\Delta\mathcal{E}^t = S^t \times B_A \times B_V \quad (16)$$

If the future average power intensity (\bar{P}) is known, then estimated DTE is given by:

$$\widehat{DTE} = \frac{\Delta\mathcal{E}^t}{\bar{P}} \quad (17)$$

The DTE prediction based on approaches using participatory sensing data is compared with the on-board DTE meter on Nissan LEAF (also known as Guess-O-Meter), which is captured by a camera mounted over the dashboard. To compare the effectiveness of DTE prediction, the deviation between the true DTE (which is computed in an offline manner) and the estimated \widehat{DTE} is measured by:

$$\Delta DTE = DTE - \widehat{DTE} \quad (18)$$

The deviation between true DTE and the on-board DTE meter on Nissan LEAF is compared.

The results are plotted in Fig. 14 for four different trips. It is observed that all approaches using participatory sensing

Approaches	Accuracy	Path Matching	System Complexity
SPM	High	Required	High
DHM	High	No	Low
MF	High	Required	Medium
Adj.	High	Required	Low
Avg.	Low	No	Low

TABLE IV: A summary of strengths and weaknesses of various approaches.

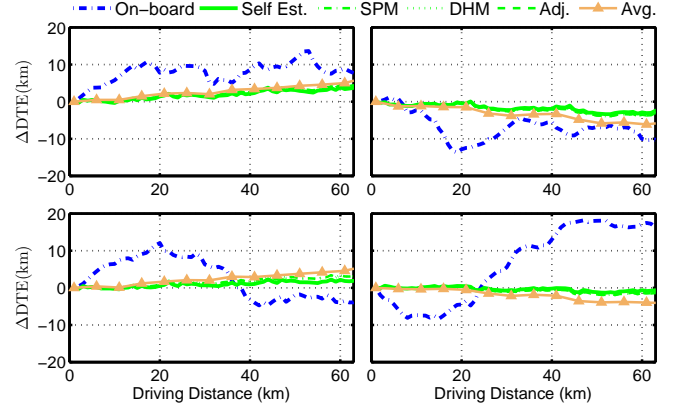


Fig. 14: Deviations of DTE prediction for various approaches.

data can significantly outperform the one provided by on-board DTE provided by the on-board DTE meter on Nissan LEAF. Notably, SPM, DHM and Adj. perform very close to benchmark Self Est., whereas Avg. gives relatively inferior performance. In summary, DHM consistently provides good accuracy with low system complexity.

VII. CONCLUSION

In this paper, the methodology for utilizing participatory sensing data for personalized prediction of driving energy consumption was investigated. Several approaches were studied and compared, including: (1) comparison with the average using personalized adjustment, (2) two similarity matching approaches based on driver/vehicle/environment-dependent factors using speed profile matching and driving habit matching, and (3) a collaborative filtering approach that uses matrix factorization. Our empirical evaluations show that participatory sensing data can significantly improve prediction accuracy. Among all approaches, similarity matching approach based on driving habits provides good accuracy (as compared to a benchmark of self-estimation using one's own driving data) with low system complexity. To evaluate the effectiveness of our approaches, a case study of DTE prediction for EVs is conducted based on the participatory sensing data. Similarity matching based on driving habit provides a practical solution to DTE prediction for EVs, which significantly outperforms the on-board DTE meter on Nissan LEAF.

Despite the promising results given by our study, there are several practical issues and limitations to be recognized:

- *Road Grade*: This is the elevation of roads. There are public mapping APIs to provide road elevation data. This can be added to the future energy consumption model.

- *Weather and Traffic*: Our study assumes mild weather and traffic conditions. Our energy consumption model can be extended to incorporate additional parameters to capture the impacts of weather in the vehicle model (e.g., weather types and route conditions). The vehicle speed from participatory sensing data can reflect the traffic condition to a certain extent, however, the data would need to be updated more frequently. Wind speed and road surface conditions also affect driving energy consumption, but are more difficult to measure.
- *More Vehicle Information*: The weight of vehicle and tire pressures of vehicle would also introduce the error to the system, the error can be minimized obtaining such information from vehicle API (e.g., wireless tire pressure indicator).
- *Change of Vehicle State*: The engine/gearbox efficiencies or battery efficiencies change over time, due to aging of vehicle or system upgrades. The energy consumption model can adapt to these changes using new data to derive new coefficients.

A study will be conducted in future work to provide more comprehensive insights in diverse practical settings, addressing the preceding issues.

REFERENCES

- [1] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, R. A. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, and G.-S. Ahn, "The rise of people-centric sensing," *IEEE Internet Computing*, vol. 12, no. 4, pp. 12–21, 2008.
- [2] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "GreenGPS: a participatory sensing fuel-efficient maps application," in *ACM MobiSys*, 2010.
- [3] D. Karbowski, S. Pagerit, and A. Calkins, "Energy consumption prediction of a vehicle along specified real-world trip," in *IEEE EVS*, 2012.
- [4] S. Grubwinkler and M. Lienkamp, "A modular and dynamic approach to predict the energy consumption of electric vehicles," in *Conf. on Future Automotive Technology*, 2013.
- [5] R. Gemulla, P. J. Haas, E. Nijkamp, and Y. Sismanis, "Large-scale matrix factorization with distributed stochastic gradient descent," in *ACM SIGKDD*, 2011.
- [6] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *J. of Computer*, vol. 42, pp. 30–37, 2009.
- [7] K. Kraschl-Hirschmann and M. Fellendorf, "Estimating energy consumption for routing algorithms," in *IEEE IV*, 2012.
- [8] E. Kim, J. Lee, and K. G. Shin, "Real-time prediction of battery power requirements for electric vehicles," in *IEEE/ACM ICCPS*, 2013.
- [9] A. Cappiello, I. Chabini, E. K. Nam, A. Lue, and M. A. Zed, "A statistical model of vehicle emissions and fuel consumption," in *IEEE ITSC*, 2002.
- [10] J. A. Oliva, C. Weihrauch, and T. Bertram, "A model-based approach for predicting the remaining driving range in electric vehicles," in *IEEE Prognostics and Health Management*, 2013.
- [11] K. Boriboonsomsin and M. J. Barth, "Impacts of road grade on fuel consumption and carbon dioxide emissions evidenced by use of advanced navigation systems," *J. of the Transportation Research Board*, vol. 2139, pp. 21–30, 2009.
- [12] Q. Yang, K. Boriboonsomsin, and M. Barth, "Arterial roadway energy/emissions estimation using modal-based trajectory reconstruction," in *IEEE ITSC*, 2011.
- [13] E. Wilhelm, J. Siegel, S. Mayer, L. Sadamori, S. Dsouza, C.-K. Chau, and S. Sarma, "Cloudthink: A scalable secure platform for mirroring transportation systems in the cloud," *Transport*, vol. 30, no. 3, 2015.
- [14] S. Dornbush and A. Joshi, "Streetsmart traffic: Discovering and disseminating automobile congestion using VANET," in *IEEE VTC*, 2007.
- [15] A. L. Oehlerking, "StreetSmart: modeling vehicle fuel consumption with mobile phone sensor data through a participatory sensing framework," in *Technical report*. MIT, 2011.
- [16] X. Hu, V. Leung, K. G. Li, E. Kong, H. Zhang, N. S. Surendrakumar, and P. TalebiFard, "Social drive: a crowdsourcing-based vehicular social networking system for green transportation," in *ACM Intl. Symp. on Design and analysis of intelligent vehicular networks and apps*, 2013.
- [17] C.-M. Tseng, C.-K. Chau, S. Dsouza, and E. Wilhelm, "A participatory sensing approach for personalized distance-to-empty prediction and green telematics," in *ACM E-energy*, 2015.
- [18] C.-M. Tseng, S. Dsouza, and C.-K. Chau, "A social approach for predicting distance-to-empty in vehicles," in *ACM International Conference on Future Energy Systems (e-Energy)*, 2014.
- [19] M. J. Burke, N. Sarafopoulos, and V. Q. To, "Electronic system and method for calculating distance to empty for motorized vehicles," 1994, US Patent 5,301,113.
- [20] L. Rodgers, E. Wilhelm, and D. Frey, "Conventional and novel methods for estimating an electric vehicle's distance to empty," in *ASME Intl. Conf. on Advanced Vehicle Technologies*, 2013.
- [21] A. Bolovinou, I. Bakas, A. Amditis, F. Mastrandrea, and W. Vinciotti, "Online prediction of an electric vehicle remaining range based on regression analysis," in *IEEE IEVC*, 2014.
- [22] H. Yu, F. Tseng, and R. McGee, "Driving pattern identification for ev range estimation," in *IEEE IEVC*, 2012.
- [23] K. Boriboonsomsin, M. J. Barth, W. Zhu, and A. Vu, "Eco-routing navigation system based on multisource historical and real-time traffic information," *IEEE Trans. on ITS*, vol. 13, no. 4, pp. 1694–1704, 2012.
- [24] S. Grubwinkler, T. Brunner, and M. Lienkamp, "Range prediction for evs via crowd-sourcing," in *IEEE VPPC*, 2013.
- [25] M. Sachenbacher, M. Leucker, A. Artmeier, and J. Haselmayr, "Efficient energy-optimal routing for electric vehicles," in *AAAI*, 2011.
- [26] S. Grubwinkler, M. Kugler, and M. Lienkamp, "A system for cloud-based deviation prediction of propulsion energy consumption for evs," in *IEEE ICVES*, 2013.
- [27] C.-M. Tseng and C.-K. Chau, "On the privacy of crowd-sourced data collection for distance-to-empty prediction and eco-routing," in *ACM Workshop on Electric Vehicle Systems, Data and Applications (EV-Sys)*, 2016.
- [28] C.-K. Chau, K. M. Elbassioni, and C.-M. Tseng, "Fuel minimization of plug-in hybrid electric vehicles by optimizing drive mode selection," in *ACM E-energy*, 2016.
- [29] C.-K. Chau, K. Elbassioni, and C.-M. Tseng, "Drive mode optimization and path planning for plug-in hybrid electric vehicles," Masdar Institute, Tech. Rep., 2016.
- [30] E. Ericsson, "Independent driving pattern factors and their influence on fuel-use and exhaust emission factor," *J. of Transportation Research*, vol. 6, no. 5, pp. 325–345, 2001.
- [31] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *J. of Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007.
- [32] "MAP- and MAF-Based Air/Fuel Flow Calculator," http://www.lightner.net/obd2guru/IMAP_AFcalc.html.
- [33] C.-M. Tseng, W. Zhou, M. A. Hashmi, C.-K. Chau, S. G. Song, and E. Wilhelm, "Data extraction from electric vehicles through OBD and application of carbon footprint evaluation," in *ACM Workshop on Electric Vehicle Systems, Data and Applications (EV-Sys)*, 2016.
- [34] H. Bozdogan, "Model selection and akaike's information criterion (aic): The general theory and its analytical extensions," *J. of Psychometrika*, vol. 52, pp. 345–370, 1987.
- [35] U. E. P. Agency, *New Fuel Economy and Environment Labels for a New Generation of Vehicles*. US Environmental Protection Agency, 2011.